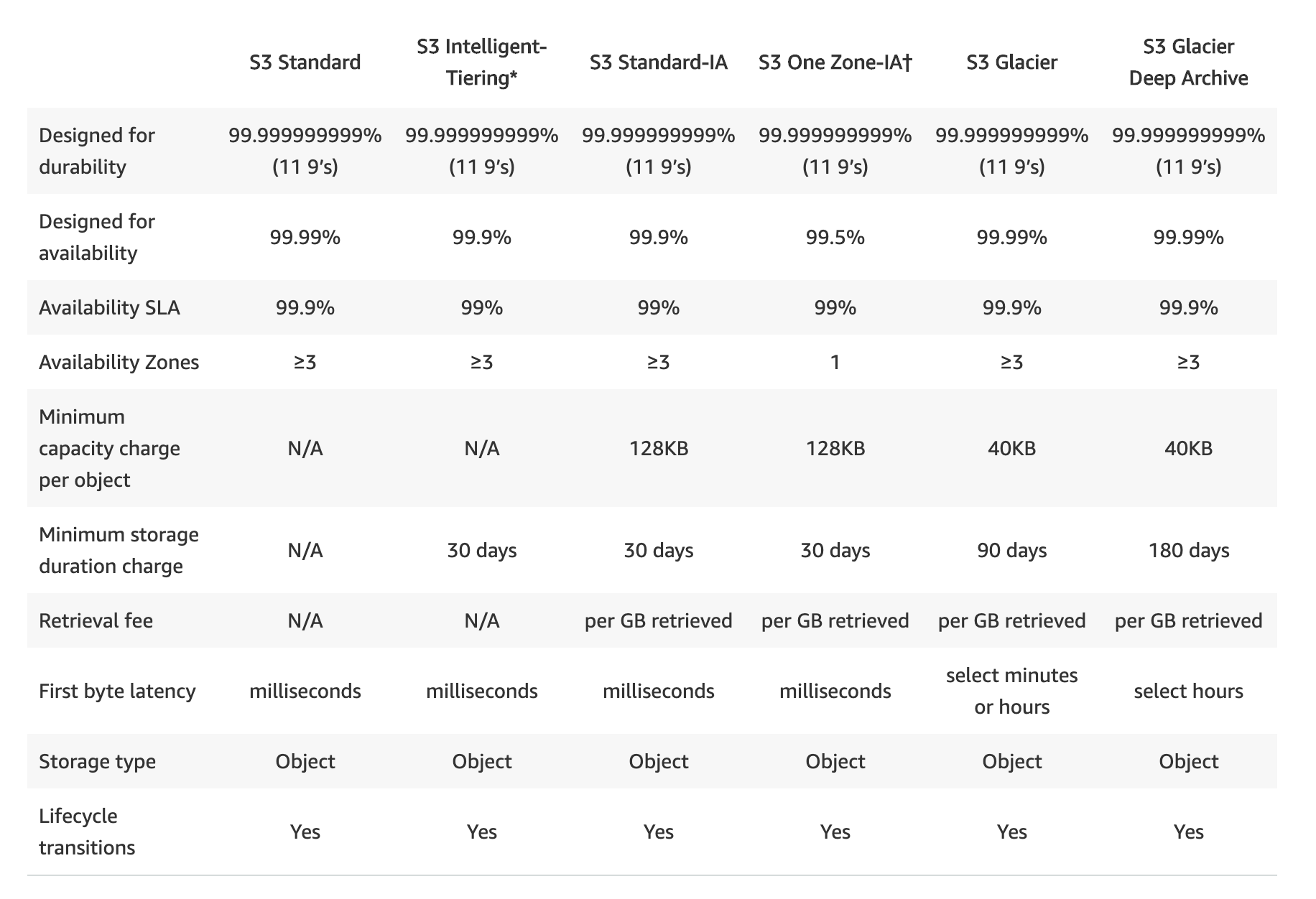
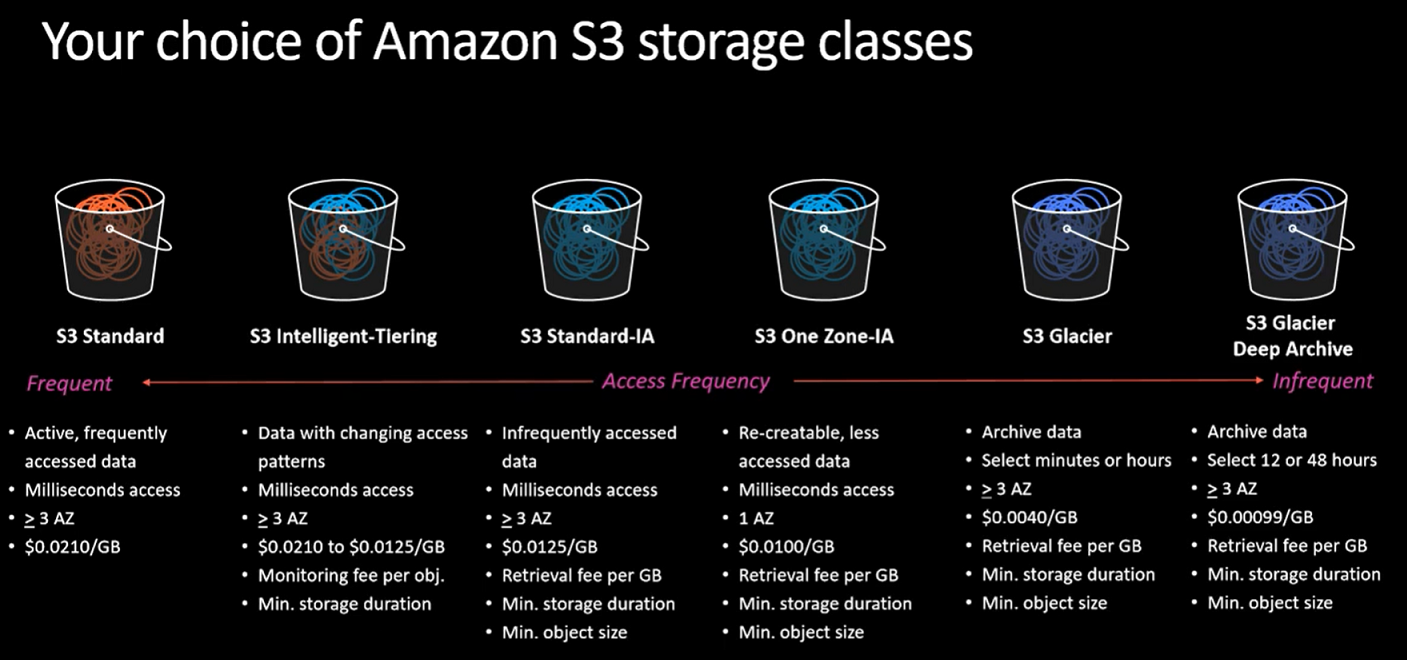
AWS S3 Storage Classes Overview

# 

# AWS S3 Storage Classes Overview

* Amazon S3 storage classes are designed to **sustain the concurrent loss of data in one or two facilities**
* S3 storage classes allows lifecycle management for automatic migration of objects for cost savings
* All S3 storage classes support SSL encryption of data in transit and data encryption at rest
* S3 also regularly verifies the integrity of your data using checksums and provides auto healing capability





## S3 Standard

* STANDARD is the **default storage class**, if none specified during upload
* Low latency and high throughput performance
* Designed for **99.999999999% i.e. 11 9’s Durability** of objects across AZs
* Designed for 99.99% availability over a given year
* Resilient against events that impact an entire Availability Zone and is designed to **sustain the loss of data in a two facilities**
* Ideal for performance-sensitive use cases and frequently accessed data
* S3 Standard is appropriate for a wide variety of use cases, including cloud applications, dynamic websites, content distribution, mobile and gaming applications, and big data analytics.

## S3 Intelligent Tiering (S3 Intelligent-Tiering)

* S3 Intelligent Tiering storage class is designed to **optimize storage costs** by automatically moving data to the most cost-effective storage access tier, without performance impact or operational overhead.
* Delivers automatic cost savings by moving data on a granular object level between two access tiers
  + one tier that is optimized for frequent access and
  + another lower-cost tier that is optimized for infrequently accessed data.
* a frequent access tier and a lower-cost infrequent access tier, when access patterns change.
* **Ideal to optimize storage costs automatically for long-lived data when access patterns are unknown or unpredictable.**
* For a small monthly monitoring and automation fee per object, S3 monitors access patterns of the objects and moves objects that have not been accessed for 30 consecutive days to the infrequent access tier.
* **There are no separate retrieval fees when using the Intelligent Tiering storage class. If an object in the infrequent access tier is accessed, it is automatically moved back to the frequent access tier.**
* **No additional fees apply when objects are moved between access tiers**
* Suitable for larger objects greater than **128 KB** (smaller objects are charged for 128 KB only) kept for at least 30 days (charged for minimum 30 days)
* Same low latency and high throughput performance of S3 Standard
* Designed for **99.999999999% i.e. 11 9’s Durability** of objects across AZs
* Designed for **99.9% availability** over a given year

## S3 Standard-Infrequent Access (S3 Standard-IA)

* S3 Standard-Infrequent Access storage class is optimized for long-lived and **less frequently accessed** data. for e.g. for  backups and older data where access is limited, but the use case still demands high performance
* Ideal for use for the primary or only copy of data that can’t be recreated.
* Data stored redundantly across multiple geographically separated AZs and are resilient to the **loss of an Availability Zone**.
* offers greater availability and resiliency than the ONEZONE\_IA class.
* Objects are available for real-time access.
* Suitable for larger objects greater than 128 KB (smaller objects are charged for 128 KB only) kept for at least 30 days (charged for minimum 30 days)
* Same low latency and high throughput performance of Standard
* Designed for **99.999999999% i.e. 11 9’s Durability** of objects across AZs
* Designed for **99.9% availability** over a given year
* S3 charges a retrieval fee for these objects, so they are most suitable for infrequently accessed data.

## S3 One Zone-Infrequent Access (S3 One Zone-IA)

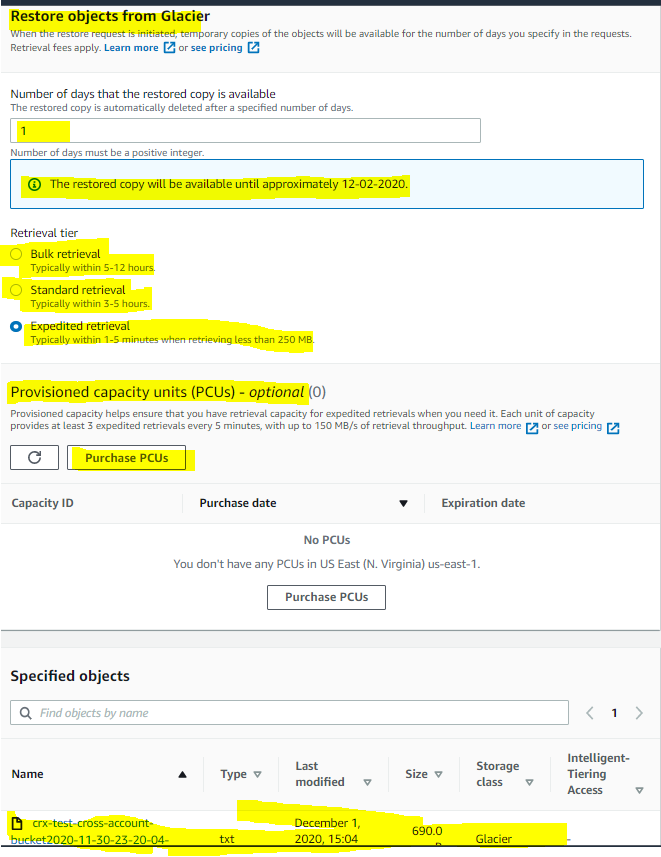
* S3 One Zone-Infrequent Access storage classes are designed for long-lived and infrequently accessed data, but available for millisecond access (similar to the STANDARD and STANDARD\_IA storage class).
* Ideal when the data can be recreated if the AZ fails, and for object replicas when setting cross-region replication (CRR).
* Objects are available for real-time access.
* Suitable for larger objects greater than 128 KB (smaller objects are charged for 128 KB only) kept for at least 30 days (charged for minimum 30 days)
* Stores the object data in **only one AZ**, which makes it less expensive than Standard-Infrequent Access
* Data is **not resilient** to the physical loss of the AZ resulting from disasters, such as earthquakes and floods.
* One Zone-Infrequent Access storage class is as durable as Standard-Infrequent Access, but it is less available and less resilient.
* Designed for **99.999999999% i.e. 11 9’s Durability** of objects in a single AZ
* Designed for **99.5% availability** over a given year
* S3 charges a retrieval fee for these objects, so they are most suitable for infrequently accessed data.

## Reduced Redundancy Storage – RRS

* **NOTE** – AWS recommends not to use this storage class. The STANDARD storage class is more cost effective.
* Reduced Redundancy Storage (RRS) storage class is designed for noncritical, reproducible data stored at lower levels of redundancy than the STANDARD storage class, which reduces storage costs
* Designed for durability of 99.99% of objects
* Designed for 99.99% availability over a given year
* Lower level of redundancy results in less durability and availability
* RRS stores objects on multiple devices across multiple facilities, providing 400 times the durability of a typical disk drive,
* RRS does not replicate objects as many times as S3 standard storage and is designed to sustain the loss of data in a single facility.
* If an RRS object is lost, S3 returns a 405 error on requests made to that object
* S3 can send an event notification, configured on the bucket, to alert a user or start a workflow when it detects that an RRS object is lost which can be used to replace the lost object

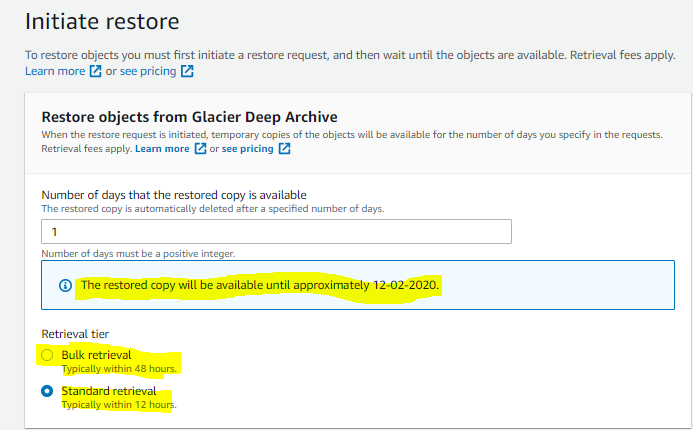
## S3 Glacier

* GLACIER storage class is suitable for **low cost data archiving** where data access is infrequent and **retrieval time of minutes to hours**  is acceptable.
* Storage class has a **minimum storage duration period of 90 days**
* Provides **configurable retrieval times**, from minutes to hours
* GLACIER storage class uses the very low-cost Glacier storage service, but the objects in this storage class are still managed through S3
* For accessing GLACIER objects,
  + object must be restored which can taken anywhere between minutes to hours
  + objects are only available for the time period (number of days) specified during the restoration request
  + object’s storage class remains GLACIER
  + charges are levied for both the archive (GLACIER rate) and the copy restored temporarily
* Vault Lock feature enforces compliance via a lockable policy
* Offers the same durability and resiliency as the STANDARD storage class
* Designed for **99.999999999% i.e. 11 9’s Durability** of objects across AZs
* Designed for **99.9% availability** over a given year

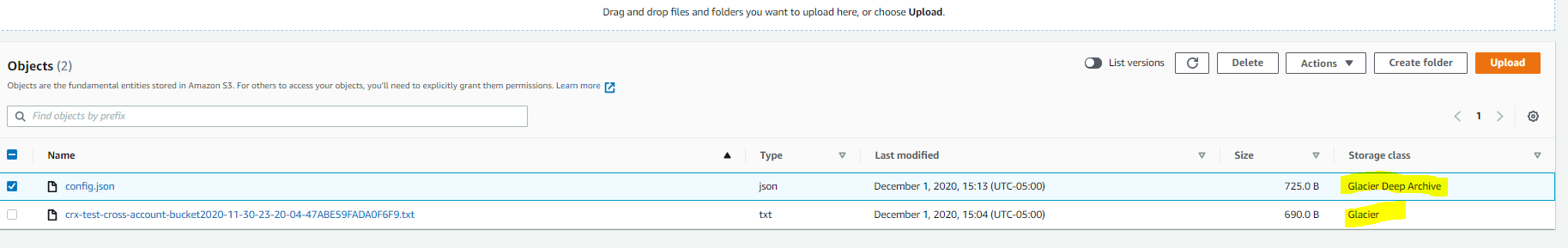


## S3 Glacier Deep Archive

* Glacier Deep Archive storage class provides **lowest cost data archiving** where data access is infrequent and retrieval time of hours  is acceptable.
* Has a **minimum storage duration period of 180 days** and can be accessed in at a default retrieval time of 12 hours.
* Supports long-term retention and digital preservation for data that may be accessed once or twice in a year
* Designed for **99.999999999% i.e. 11 9’s Durability** of objects across AZs
* Designed for **99.9% availability** over a given year
* DEEP\_ARCHIVE retrieval costs can be reduced by using bulk retrieval, which returns data within 48 hours.



* Glacier & Glacier Deep Archive storage class objects would not be accessible until they restore process is complete.
* Once restored period is expired, they will not be accessible again.



## AWS Certification Exam Practice Questions

1. What does RRS stand for when talking about S3?
   1. Redundancy Removal System
   2. Relational Rights Storage
   3. Regional Rights Standard
   4. **Reduced Redundancy Storage**
2. What is the durability of S3 RRS?
   1. **99.99%**
   2. 99.95%
   3. 99.995%
   4. 99.999999999%
3. What is the Reduced Redundancy option in Amazon S3?
   1. **Less redundancy for a lower cost**
   2. It doesn’t exist in Amazon S3, but in Amazon EBS.
   3. It allows you to destroy any copy of your files outside a specific jurisdiction.
   4. It doesn’t exist at all
4. An application is generating a log file every 5 minutes. The log file is not critical but may be required only for verification in case of some major issue. The file should be accessible over the internet whenever required. Which of the below mentioned options is a best possible storage solution for it?
   1. AWS S3
   2. AWS Glacier
   3. AWS RDS
   4. **AWS S3 RRS**(Reduced Redundancy Storage (RRS) is an Amazon S3 storage option that enables customers to store noncritical, reproducible data at lower levels of redundancy than Amazon S3’s standard storage. RRS is designed to sustain the loss of data in a single facility.)
5. A user has moved an object to Glacier using the life cycle rules. The user requests to restore the archive after 6 months. When the restore request is completed the user accesses that archive. Which of the below mentioned statements is not true in this condition?
   1. The archive will be available as an object for the duration specified by the user during the restoration request
   2. **The restored object’s storage class will be RRS**(After the object is restored the storage class still remains GLACIER. [Read more](http://docs.aws.amazon.com/AmazonS3/latest/dev/restoring-objects.html))
   3. The user can modify the restoration period only by issuing a new restore request with the updated period
   4. The user needs to pay storage for both RRS (restored) and Glacier (Archive) Rates
6. Your department creates regular analytics reports from your company’s log files. All log data is collected in Amazon S3 and processed by daily Amazon Elastic Map Reduce (EMR) jobs that generate daily PDF reports and aggregated tables in CSV format for an Amazon Redshift data warehouse. Your CFO requests that you optimize the cost structure for this system. Which of the following alternatives will lower costs without compromising average performance of the system or data integrity for the raw data? **[PROFESSIONAL]**
   1. Use reduced redundancy storage (RRS) for PDF and CSV data in Amazon S3. Add Spot instances to Amazon EMR jobs. Use Reserved Instances for Amazon Redshift. (Spot instances impacts performance)
   2. **Use reduced redundancy storage (RRS) for all data in S3. Use a combination of Spot instances and Reserved Instances for Amazon EMR jobs. Use Reserved instances for Amazon Redshift**(Combination of the Spot and reserved with guarantee performance and help reduce cost. Also, RRS would reduce cost and guarantee data integrity, which is different from data durability )
   3. Use reduced redundancy storage (RRS) for all data in Amazon S3. Add Spot Instances to Amazon EMR jobs. Use Reserved Instances for Amazon Redshift (Spot instances impacts performance)
   4. Use reduced redundancy storage (RRS) for PDF and CSV data in S3. Add Spot Instances to EMR jobs. Use Spot Instances for Amazon Redshift. (Spot instances impacts performance)
7. Which of the below mentioned options can be a good use case for storing content in AWS RRS?
   1. Storing mission critical data Files
   2. Storing infrequently used log files
   3. Storing a video file which is not reproducible
   4. **Storing image thumbnails**
8. A newspaper organization has an on-premises application which allows the public to search its back catalogue and retrieve individual newspaper pages via a website written in Java. They have scanned the old newspapers into JPEGs (approx. 17TB) and used Optical Character Recognition (OCR) to populate a commercial search product. The hosting platform and software is now end of life and the organization wants to migrate its archive to AWS and produce a cost efficient architecture and still be designed for availability and durability. Which is the most appropriate? **[PROFESSIONAL]**
   1. Use S3 with reduced redundancy to store and serve the scanned files, install the commercial search application on EC2 Instances and configure with auto-scaling and an Elastic Load Balancer. (RRS impacts durability and commercial search would add to cost)
   2. Model the environment using CloudFormation. Use an EC2 instance running Apache webserver and an open source search application, stripe multiple standard EBS volumes together to store the JPEGs and search index. (Using EBS is not cost effective for storing files)
   3. **Use S3 with standard redundancy to store and serve the scanned files, use CloudSearch for query processing, and use Elastic Beanstalk to host the website across multiple availability zones.**(Standard S3 and Elastic Beanstalk provides availability and durability, Standard S3 and CloudSearch provides cost effective storage and search)
   4. Use a single-AZ RDS MySQL instance to store the search index and the JPEG images use an EC2 instance to serve the website and translate user queries into SQL. (RDS is not ideal and cost effective to store files, Single AZ impacts availability)
   5. Use a CloudFront download distribution to serve the JPEGs to the end users and Install the current commercial search product, along with a Java Container for the website on EC2 instances and use Route53 with DNS round-robin. (CloudFront needs a source and using commercial search product is not cost effective)
9. A research scientist is planning for the one-time launch of an Elastic MapReduce cluster and is encouraged by her manager to minimize the costs. The cluster is designed to ingest 200TB of genomics data with a total of 100 Amazon EC2 instances and is expected to run for around four hours. The resulting data set must be stored temporarily until archived into an Amazon RDS Oracle instance. Which option will help save the most money while meeting requirements? **[PROFESSIONAL]**
   1. **Store ingest and output files in Amazon S3. Deploy on-demand for the master and core nodes and spot for the task nodes.**
   2. Optimize by deploying a combination of on-demand, RI and spot-pricing models for the master, core and task nodes. Store ingest and output files in Amazon S3 with a lifecycle policy that archives them to Amazon Glacier. (Master and Core must be RI or On Demand. Cannot be Spot)
   3. Store the ingest files in Amazon S3 RRS and store the output files in S3. Deploy Reserved Instances for the master and core nodes and on-demand for the task nodes. (Need better durability for ingest file. Spot instances can be used for task nodes for cost saving.)
   4. Deploy on-demand master, core and task nodes and store ingest and output files in Amazon S3 RRS (Input must be in S3 standard)