**Content**

[1. Concepts 2](#_Toc69133370)

[1. What is Data Analytics 2](#_Toc69133371)

[2. OLTP vs OLAP 2](#_Toc69133372)

[2. S3 3](#_Toc69133373)

[1. Global Accelerator 3](#_Toc69133374)

[2. Transfer accelerator 3](#_Toc69133375)

[3. S3 Consistency 3](#_Toc69133376)

[4. Multi-part upload 4](#_Toc69133377)

[5. S3 and Boto3 5](#_Toc69133378)

[6. S3 storage classes 6](#_Toc69133379)

[7. Glacier retrieval methods 7](#_Toc69133380)

[8. S3 Security 7](#_Toc69133381)

[3. Glue 9](#_Toc69133382)

[4. Kinesis 10](#_Toc69133383)

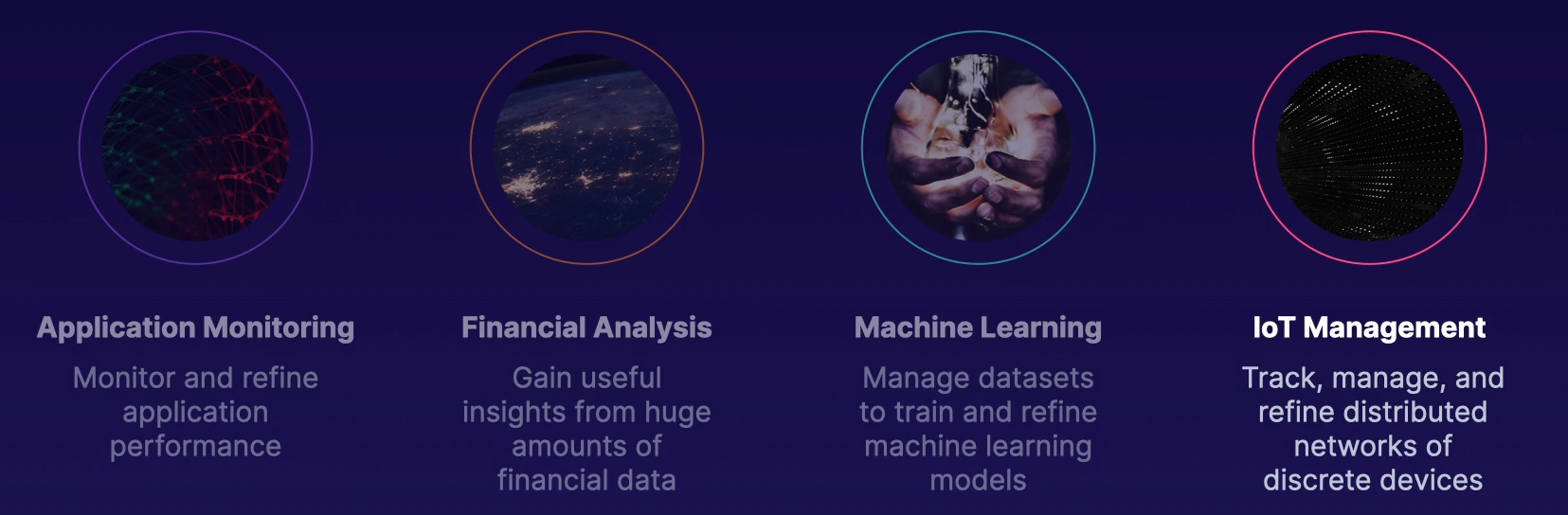
[5. Redshift 11](#_Toc69133384)

1. Concepts
   1. What is Data Analytics

Data Analytics workflow:

* define
* data collection
* data preparation
* data analysis
* data interpretation
* discovery

Possible applications:



* 1. OLTP vs OLAP

1. S3
   1. Global Accelerator
   2. Transfer accelerator

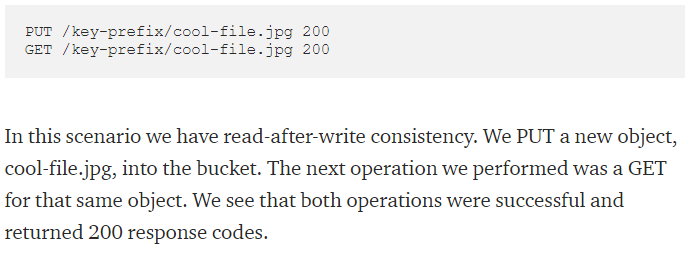
* uses the CF edge locations for data ingestion. Enabled per bucket, additional cost. Transfer acceleration endpoints:

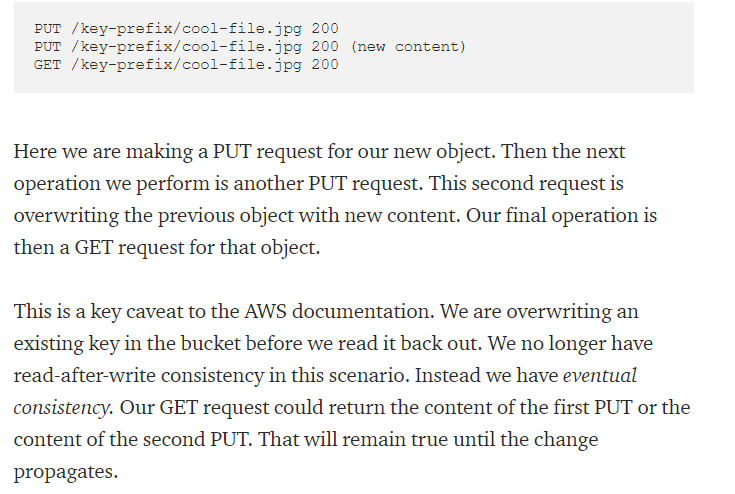
bucketname.s3-accelerate.amazonaws.com

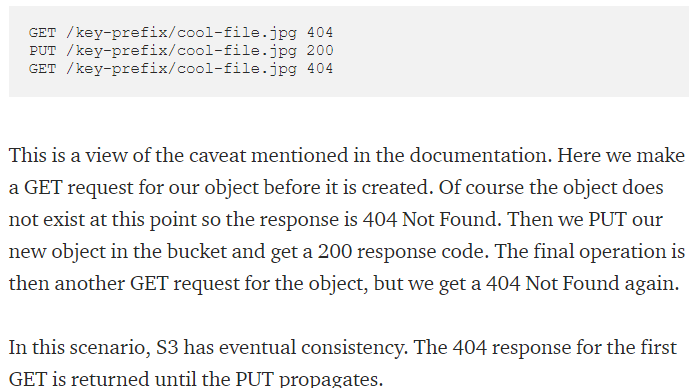
S3 + cloudfront

* 1. S3 Consistency

Amazon S3 delivers strong read-after-write consistency. After a successful write of a new object, or an overwrite or delete of an existing object, any subsequent read request immediately receives the latest version of the object. S3 also provides strong consistency for list operations, so after a write, you can immediately perform a listing of the objects in a bucket with any changes reflected



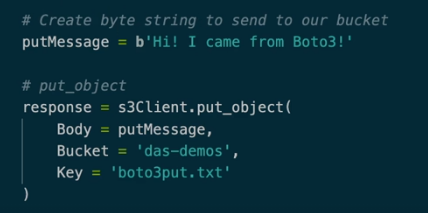




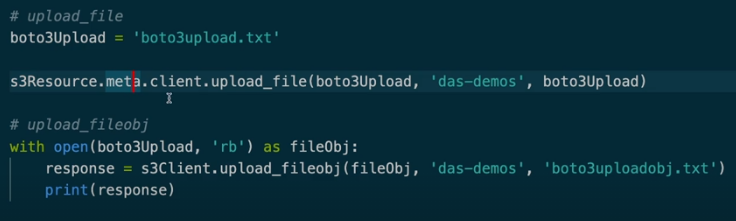
* 1. Multi-part upload

1. CreateMultipartUpload API Call, returns bucket, key and uploadID.
2. UploadPart API Call, provide bucket, key, part number and UploadID. Returns Etag. you upload individual parts of the file
3. CompleteMultipartUpload api call.Provide bucket, key, updateId and Etags

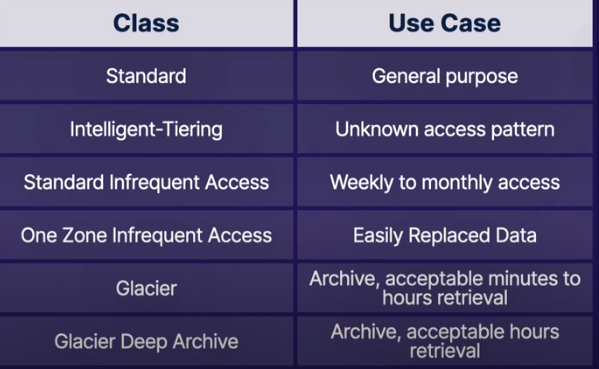
* Multipart uploads can be made up of up to 10k parts.
* You can update a part – even retry and update if changed
* Auto-abort for multipart uploads taking too long
* 100 MB+ multipart upload
* All parts except the funal aprt must be at least 5 MB (between 5-100 MB)
  1. S3 and Boto3

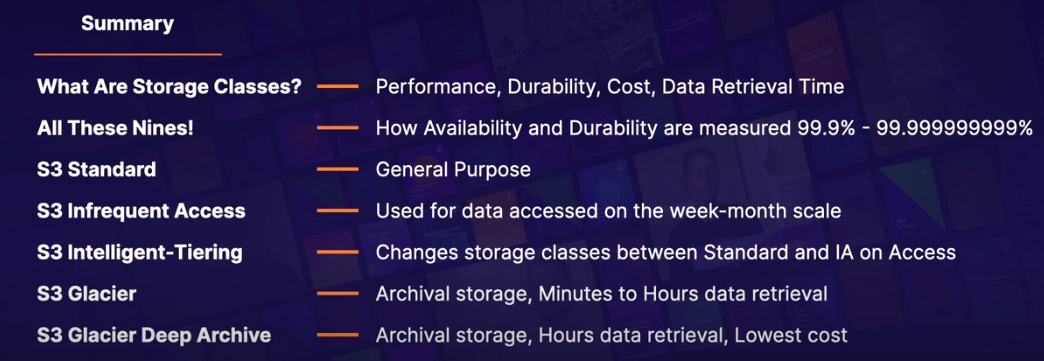






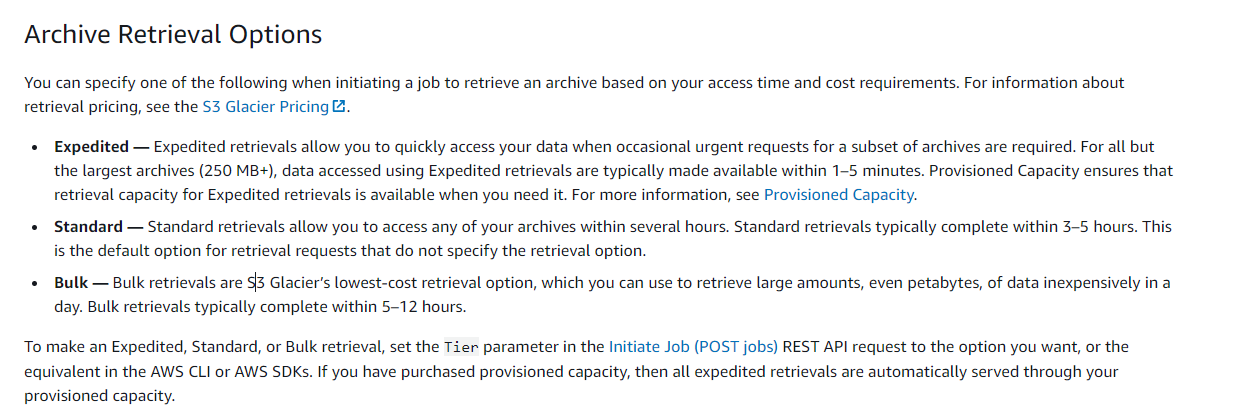
* 1. S3 storage classes

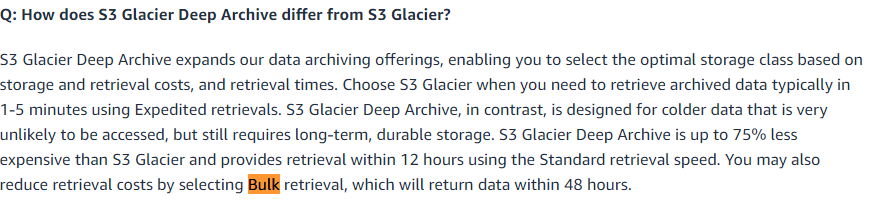




* In the lifecycle rules the time intervals are not additive
* In the licecylce rules, you can move objects based on „Prefix“ and „Tag“. The transition itself define via „Transitions“.
* In the lifecycle rules, you can set „Expirations“ in „Days“ when to remove the data.\
* There is cost for the transition between storage types. If you have lots and lots of small files this can be expensive. To solve, reduce the scope via prefixes and tags.
  1. Glacier retrieval methods







* 1. S3 Security

S3 security features

* Access Analyzer for S3
* Amazon S3 Server Access Logging
* Bucket Policy
* Bucket ACL
* Cross-region replication

To enable this versioning must be enabled.

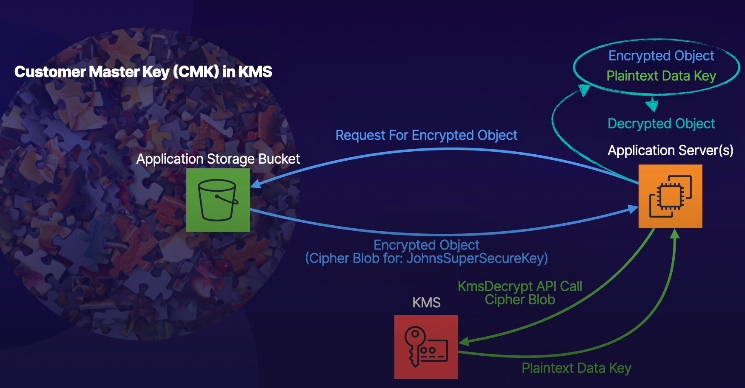
* MFA for DELETE
* Object ACL
* Object Lock
* Versioning
* CW Alarms
* Cloudtrail logs
* IAM, SCP, KMS
* VPC Endpoints

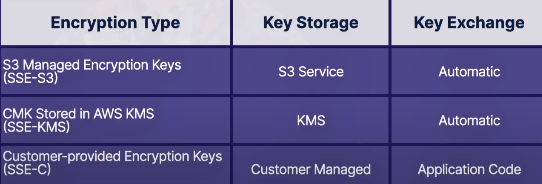
Security

* In flight security
  + Mandatory Client TLS Support
    - All S3 clients must support at least TLS1.0 thoguh TLS 1.2+ is recommended
    - Additionally clients must support PFS (Perfect Forwards Secrecy) cipher suites: DHE (Ephemeral Diffie-Hellman), ECDHE (Elliptic Curve Diffie-Hellman Ephemeral)
  + VPC Endpoints – so the data does not traverse thoguh the public internet. Use vpns to access the bucket later on from outside. (AWS Site-to-Site VPN, Direct Connect)
* At rest
  + Client side encryption
    - Customer Master Key (CMK) in KMS
    - Use a master key that you store within your application.

Available in AWS SDK for .NET, Go, PHP, Java, Ruby, C++, Java

* + Server side encryption
    - S3 Managed Encryption Keys (SSE-S3)
    - Customer-provided Encryption Keys (SSE-C)
    - CMK Stored in AWS KMS (SSE-KMS)





1. Glue
2. Kinesis
3. Redshift